BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e39215
doi: 10.3897/biss.3.39215

OPEN ACCESS

Conference Abstract

From Raw Biodiversity Data to Indicators, Boosting Products Creation, Integration and Dissemination: French BON FAIR initiatives and related informatics solutions

Yvan Le Bras[‡], Aurélie Delavaud[§], Dominique Pelletier[|], Jean-Baptiste Mihoub[¶]

‡ French Museum of Natural History, Concarneau, France
§ French Foundation for Research on Biodiversity, Paris, France
| Ifremer, Nantes, France
¶ Sorbonne University, Paris, France

Corresponding author: Yvan Le Bras (yvan.le-bras@mnhn.fr)

Received: 18 Aug 2019 | Published: 20 Aug 2019

Citation: Le Bras Y, Delavaud A, Pelletier D, Mihoub J-B (2019) From Raw Biodiversity Data to Indicators, Boosting Products Creation, Integration and Dissemination: French BON FAIR initiatives and related informatics solutions. Biodiversity Information Science and Standards 3: e39215. https://doi.org/10.3897/biss.3.39215

Abstract

Most biodiversity research aims at understanding the states and dynamics of biodiversity and ecosystems. To do so, biodiversity research increasingly relies on the use of digital products and services such as raw data archiving systems (e.g. structured databases or data repositories), ready-to-use datasets (e.g. cleaned and harmonized files with normalized measurements or computed trends) as well as associated analytical tools (e.g. model scripts in Github). Several world-wide initiatives facilitate the open access to biodiversity data, such as the Global Biodiversity Information Facility (GBIF) or GenBank, Predicts etc. Although these pave the way towards major advances in biodiversity research, they also typically deliver data products that are sometimes poorly informative as they fail to capture the genuine ecological information they intend to grasp. In other words, access to ready-to-use aggregated data products may sacrifice ecological relevance for data harmonization, resulting in over-simplified, ill-advised standard formats. This is singularly true when the main challenge is to match complementary data (large diversity of measured variables, integration of different levels of life organizations etc.) collected with

© Le Bras Y et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

different requirements and scattered in multiple databases. Improving access to raw data, and meaningful detailed metadata and analytical tools associated with standardized workflows is critical to maintain and maximize the generic relevance of ecological data. Consequently, advancing the design of digital products and services is essential for interoperability while also enhancing reproducibility and transparency in biodiversity research. To go further, a minimal common framework organizing biodiversity observation and data organization is needed. In this regard, the Essential Biodiversity Variable (EBV) concept might be a powerful way to boost progress toward this goal as well as to connect research communities worldwide.

As a national Biodiversity Observation Network (BON) node, the French BON is currently embodied by a national research e-infrastructure called "Pôle national de données de biodiversité" (PNDB, formerly ECOSCOPE), aimed at simultaneously empowering the quality of scientific activities and promoting networking within the scientific community at a national level. Through the PNDB, the French BON is working on developing biodiversity data workflows oriented toward end services and products, both from and for a research perspective. More precisely, the two pillars of the PNDB are a metadata portal and a workflow-oriented web platform dedicated to the access of biodiversity data and associated analytical tools (Galaxy-E). After four years of experience, we are now going deeper into metadata specification, dataset descriptions and data structuring through the extensive use of Ecological Metadata Language (EML) as a pivot format. Moreover, we evaluate the relevance of existing tools such as Metacat/Morpho and DEIMS-SDR (Dynamic Ecological Information Management System - Site and dataset registry) in order to ensure a link with other initiatives like Environmental Data Initiative, DataOne and Long-Term Ecological Research related observation networks. Regarding data analysis, an open-source Galaxy-E platform was launched in 2017 as part of a project targeting the design of a citizen science observation system in France (“65 Millions d'observateurs”).

Here, we propose to showcase ongoing French activities towards global challenges related to biodiversity information and knowledge dissemination. We particularly emphasize our focus on embracing the FAIR (findable, accessible, interoperable and reusable) data principles Wilkinson et al. 2016 across the development of the French BON e-infrastructure and the promising links we anticipate for operationalizing EBVs. Using accessible and transparent analytical tools, we present the first online platform allowing the performance of advanced yet user-friendly analyses of biodiversity data in a reproducible and shareable way using data from various data sources, such as GBIF, Atlas of Living Australia (ALA), eBIRD, iNaturalist and environmental data such as climate data.

Keywords

biodiversity data, raw data, metadata, EML, Ecological Metadata Language, indicators, EBV, Essential Biodiversity Variable, French BON, Metacat, Galaxy, Galaxy-E, GO FAIR, BiodiFAIRse, PNDB, Pôle national de données de biodiversité

Presenting author

Yvan Le Bras

Presented at

Biodiversity_Next 2019

Hosting institution

French National Museum of Natural History

References

- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1): 160018. https://doi.org/10.1038/sdata.2016.18